

日本語教育におけるコーパスの応用

ーデータ駆動型学習とその実践方法の考察ー

寺嶋 弘道

アブストラクト

コンピューター技術の発展に伴い、コーパスが構築され、それらのコーパスが言語教育で応用できるようになってきた。学習者がコーパスを使い、学習する方法はデータ駆動型学習（Data-driven-Learning）と呼ばれ、学習者がオーセンティック（authentic）な言語に大量に触れながら、学習できるという特徴がある。しかし、現在、データ駆動型学習が日本語教育で広く取り入れられているとはいえず、そういった実践例もほとんど見られない。そこで、本稿では日本語教育分野におけるデータ駆動型学習の基礎的な資料となるよう、データ駆動型学習の方法、その効果と問題点、問題の改善方法について整理し、本学におけるデータ駆動型学習の取り組みについて報告した。

キーワード：コーパス、データ駆動型学習、漢字語彙学習、コロケーション

1. はじめに

現在のコンピューター技術の発展に伴い、コンピューター上で扱うことができる大規模なテキスト（以下コーパス）の構築が進んでいる。コーパスの言語教育分野への応用としては、教材作成、教育語彙の選定などが代表的であるが、学習者がコーパスそのものを活用し、学習するという方法も提案されてきた。このような学習方法はデータ駆動型学習（Data-driven-Learning）と呼ばれ、コーパスの構築とコンコーダンサーと呼ばれる検索システムの開発が進んでいる。現在だからこそ、可能な学習方法となっている。データ駆動型学習はバーミンガム大学教授であった Tim Johns らによって提唱された学習方法で、英語教育分野において積極的に取り入れられてきたものである。Tim Johns の他にも、Chris Tribble、Mike Scott、Guy Aston、Sylvia Bernardini といった研究者が「学習者自身が探究者となる（learner as researcher）」という発想に基づき、学習者が直接コーパスを利用する方法を取り入れている（投野 2003）。

日本語教育においても英語教育の知見を応用し、データ駆動型学習を行うための自学自習用教材を開発した中條他（2009）の取り組みがあるが、データ駆動型学習というアプローチそのものが十分に知られていないのが現状である。

そこで、本稿では日本語教育分野におけるデータ駆動型学習の基礎的な資料となるよう、データ駆動型学習の方法、その効果と問題点、問題の改善方法について整理したい。そして、実際に本学で行っているデータ駆動型学習の取り組みについて報告したい。

2. データ駆動型学習の方法とその効果

2.1 学習方法

投野（2003）はコーパスを言語教育に応用する場合、直接的なアプローチと間接的なアプローチがあるとしている。間接的なアプローチとは、コーパスから得られた情報をシラバスや教材の作成に生かす方法であるのに対し、直接的なアプローチとは学習者が自らコンコーダンサーを用いる方法である。しかし、実際のデータ駆動型学習では Jones（1994）、Tribble and Jones（1997）といった英語教育分野の先行研究に代表されるように、直接的アプローチと間接的アプローチを組み合わせた方法が取られていることが多い。その方法とは教師がコーパスに基づき作成した教材を提供し、学習者がその教材のタスクに答えるためにコーパスを直接利用し、学習を行うというものである。日本語教育分野においても、中條他（2009）がコーパスから得られた情報に基づいて作成した教材を「DDL教材」とし、「対応」、「援助」、「態度」といった語を学ぶための教材を開発している。その教材には、日英パラレルコーパスから検索できるコンコーダンサーで

共起表現を探し、それを分類するタスク、空欄にされた共起表現が何かを考えるタスク等が含まれている。一方、直接的アプローチの例としてはNoguchi (2004)、岩井 (2009) のように学習者がコーパスを直接利用し、コーパスから得られた情報を英作文に生かしたものがある。日本語教育分野でも国立国語研究所によるコーパスを作文に応用するための研究プロジェクトが進行中である¹。以上の先行研究はDDL教材での学習や作文の支援といった目的でコーパスを用いた例である。しかし、学習者が辞書や文法書などとともにコーパスを使用すれば、指導者がいない場面でも自分の疑問を解決することができる (梅咲 2008) という指摘もあるように、データ駆動型学習を「学習者自らがコーパスを用いる過程で生じる学習」だと捉えれば、様々な目的でデータ駆動型学習を導入することが可能である。

データ駆動型学習の特徴は、帰納的な学習が行える点であると説明されることがある (中條 2008)。その典型的な学習は、大規模なデータから用例を検索できるというコンコーダンスーの特徴を生かし、帰納的に文法的規則性やよく使用される共起表現を発見するというものである。コンコーダンスーはソート機能で類似した表現を近くに表示できるため、より短時間での発見が期待できる。しかし、データ駆動型学習は必ず帰納的なアプローチを取らなければならないわけではない。たとえば、演繹的に学んだ内容が具体的にどのように使用されているかを検証しながら、個々の例から学ぶ方法を取り入れることも可能である。コーパスの中には演繹的に学んだ内容とは異なる場合もあるため、そのような例外は新たな知識を獲得する機会となる。

また、投野 (2003) はデータ駆動型学習には分散型 (divergent) と収斂型 (convergent) の学習方法があるとしている。分散型というのはデータを観察し、出した結論が学習者によって異なるというもので、収斂型というのは教師の側で学習者が到達する結論をあらかじめ用意し、最終的に学習者が共通理解を得ることを目指すものである。あらかじめ、結論を準備する収斂型の場合、教師が直接その結論を教えれば時間の無駄も省くことができるが、先にも述べたようにデータ駆動型学習は学習者が自らコンコーダンスーを用いる過程で学ぶことが重視されるため、そのような方法を取らないのが特徴である。

2.2 データ駆動型学習の効果

Jones (1994) はデータ駆動型学習の特徴としてオーセンティック (authentic) な言語に大量に触れられるという点を挙げている。教師が教材や授業で示す用例や辞書に含まれる用例は量に限りがあり、不自然なものもある。しかし、データ駆動型学習は、実際に文脈がある中で使用された大量の言語表現に学習者が触れながら、学習できるという効果がある。

また、コーパスはどのような表現がよく使用されるのかを示す実証的資料となるため、コーパスを辞書や文法書とともに使用すれば、非母語話者でも自分で疑問を解決できる可能性があると考えられている (梅咲 2008)。そのため、データ駆動型学習を授業で導入することにより、言語学習するうえで有効な学習ストラテジーを教えることができる。

さらに演繹的なアプローチであっても、帰納的なアプローチであっても、主体的に分析しなければならないのがデータ駆動型学習の特徴であるため、多くの言語を観察する学習環境を提供し、繰り返し観察することで学習者の言語現象に対する分析力を高められる可能性もある。学習者が自ら目標言語についての発見をしていくという意味では、データ駆動型学習は、認知心理学の第一人者であるブルーナー (1963) が提唱した「発見学習」と通じるところがあり、発見学習と同じような学習効果が期待できる可能性がある。たとえば、水越 (1975) はブルーナーが提唱した発見学習の効果として、法則を発見しようという身構えた心を育成できる、学習への動機を高められる、記憶が長く保持されるといった点を挙げている。ブルーナーが提唱した発見学習は子供への科目教育を対象にしているが、データ駆動型学習においてもそのような効果が実際にあるのか、今後明らかにしていく必要がある。

2.3 データ駆動型学習の問題点

ここではデータ駆動型学習における問題点とその改善方法を考察したい。まず、データ駆動型学習の問題点としては、コーパスの難易度の問題やコーパスに含まれるテキストのテーマといったように、コーパスそのものと関わるものがある。そのような問題を改善するために、コンコーダンスーでは様々な工夫がされている。たとえば、国立国語研究所が公開しているKOTONOAの場合は、ジャンルや期間を選択でき²、東京国際大学の川村氏他が公開しているJpwac-L2 では、

旧日本語能力試験のLevel4(4級)～Level1(1級)、さらに上級のLevel0の選択ができ、表示される難易度を変えることができる³。そのような機能を用いることで、コーパスそのものに関わる問題は軽減できると思われる。

次に考えられるのが学習者のデータ駆動型学習に対する態度である。データ駆動型学習というのは、新しい学習方法であるため、そのような学習を経験したことがない学習者が多い。多くのデータを主体的に分析する能力が求められるデータ駆動型学習を紹介しても、自分に合わないと考える学習者もいると思われる。そのため、データ駆動型学習の導入に当たってはガイダンスを設け、データ駆動型学習にどんなメリットがあるかを説明しておく必要がある。

また、学習者のコンコーダンスーの使い方についても注意が必要である。コンコーダンスーが高機能になっているため、パソコンに苦手意識を持っている学習者はコンコーダンスーの機能を説明しても、十分に理解できないおそれがある。また、コンコーダンスーで検索する場合も語の変化や文構造を意識した検索方法を習得しておかなければ、欲しい情報を正確に取り出せない。そのため、ガイダンスの中でコンコーダンスーの機能、効果的な検索方法などを説明しておく必要がある。

最後にデータ駆動型学習を行っても、必ず欲しい情報が得られるわけではないという問題もある。コーパスの規模にも限界があるため、低頻度の表現になると、欲しい情報が見つけれられないということもある。また、例外が多すぎて、共通性を発見できないということも考えられる。そのような問題を改善するためには、事前に教師がコンコーダンスーで調査し、学習項目から外すことや中條(2008)のように、発見までの過程を教材化することで自主的な発見学習を導くという方法が考えられる。

3. 日本語教育で活用できるコーパス

最初にも述べたように、近年データ駆動型学習を行うために必要なコーパスが公開されるようになり、このような学習方法が可能になっている。しかし、コーパスには著作権の問題があるため、それほど多くのコーパスが公開されているわけではない。また、コーパスが公開されたとしても、1億語規模のコーパスでは低頻度語の現象を明らかにするためには十分ではないという考え方もある(石川2009)。そのため、荻野(2008)、石川(2008)のように、身近なものとして存在しているWWWを一つの巨大なコーパスと見なす考え方、すなわちWAC(Web as Corpus)というアプローチも存在している。たとえば、Googleなどの検索エンジンを用いれば、検索した表現がWWW上にどれくらい存在するのかがヒット件数として表示される。そのようなヒット件数を手がかりに言語現象を調査することが可能なのである。

しかし、それらは研究者による言語研究という立場で述べられているもので、学習者による言語学習となると、いくつかの問題が考えられる。まず、考えられるのは、WWW上にある検索エンジンは、コンコーダンスーのようにKWIC(Key Word in Context)表示及びソートができないため⁴、言語学習には向かないという点である。コンコーダンスーを用いれば、検索した語を中心に置き、その前後にどのような共起語が用いられるかを視覚的に分析することが可能である。しかし、Google等の検索エンジンを用いて共起表現を探そうとしても、そのような機能がないため、手間がかかってしまう。また、WWWにあるテキストの難易度や質の点から考えても、WWWをコーパスとして見なし、学習者が言語学習に活用するには無理があるように思われる。やはり、データ駆動型学習のためには、理論的にサンプリングされたコーパスと専用の検索システムが必要である。

現在、日本語教育においてデータ駆動型学習への活用が期待できるのは、WEB上で公開されているKOTONOAとJPwac-L2である⁵。現在、前者は4600万語、後者は1300万語規模のコーパスから検索が可能である。以下、KOTONOAとJPwac-L2に分け、その機能を説明したい。

3.1 KOTONOAの機能

KOTONOAとは、国立国語研究所がWEB上で公開しているコンコーダンスーで、同研究所で構築している現代日本語書き言葉均衡コーパス(以下BCCWJ)から検索を行うことができる。BCCWJの特徴は、書籍、白書、Yahoo知恵袋といった多様な内容が含まれていること、そして、母集団の特性が反映されるよう、構築されていることである。現在(2010年3月)検索対象となっているのは、一般の書籍からのサンプル(8821件、約2500万語)、政府刊行白書からのサンプル(1500件、約500万語)、過去30年間の国会会議録からのサンプル(159会議、約500万語)、2005年度版の検定教科書からの

サンプル (412 件、約 100 万語)、「Yahoo!知恵袋」からのサンプル (45725 件、約 500 万語)、「Yahoo!ブログ」からのサンプル (24027 件、約 500 万語) の 6 種類のデータ合計 4600 万語である。

図 1 は KOTONOA で検索を開始する画面である。検索する際には、チェックボックス機能を用い、ジャンル及び期間を設定することができる。ジャンルは、先に述べた BCCWJ に含まれるジャンルを検索に含めるのかどうかを選択できる。さらに、各ジャンルの+をクリックすると、より細分化したジャンルが表示される。たとえば、書籍の場合であれば、NDC 番号で 0～9、無分類で選ぶことができる。一方、期間の場合、出版された期間を限定し、検索することができる。

検索された結果は図 2 のように検索語の左右に出現する表現が見やすく表示される。ソート機能で左右の表現を並び換えれば、同じ表現がどれくらいあるかを知ることができる。検索語の左側でソートをする場合は左文脈、右側の場合には、右文脈をクリックすればよい。KOTONOA は検索結果が 500 件以上の場合、一旦コーパス全体を検索したうえで、無作為に 500 件を選んで表示するようになっている。

図 1 KOTONOA の検索画面



図 2 KOTONOHA の検索結果

KOTONOHA「現代日本語書き言葉均衡コーパス」 検索デモンストレーション - Mozilla Firefox

KOTONOHA「現代日本語書き言葉均衡コーパス」

検索結果
852 件の結果が見つかりました。そのうち 500 件を表示しています。

表示番号	前文脈	検索文字列	後文脈	執筆者	生年代	性別	ジャンル	書名/出典	副題/分類	巻号	編著者等*	出版者	出版年	備考
1	して核燃料サイクルシステムの安全管理に	取り組む	必要がある。当面、以下のような所要の措置	総理府 原子力 安全委 員会			白書/安全/原子 力安全白書	原子力安全白 書_平成11年 版				大蔵省 印刷局	2000	
2	、「取り組むかどうかを検討中」を1点、「	取り組む	予定はない」を0点にして得点を試算した。	前田 正子	1960	女	書籍/3 社会科学	生涯現役時代 の雇用政策			清家篤1 編著	日本評 論社	2001	
3	手の定着の目標等、集落として今後5年間に	取り組む	べき事項や目標を定めるもの。(注)ここで				白書/農林水産/ 食料農業農村白書 (農業白書)	食料・農業・農 村白書_平成 13年度(図 録)				(財)農 林統計 協会	2002	
4	に考える、こういうこと。府県制度の改革に	取り組む	ということ、あるいは市町村制度も今の画一	片山虎 之助			国会会議員/参議 院/常任委員会	第154回国会 参議院 常任 委員会 総務 委員会					2002	
5	等により、公的部門全体の総人件費の抑制に	取り組む	こととしてあり、国家公務員の総人件費を今	二之湯 智			国会会議員/参議 院/常任委員会	第168回国会 参議院 常任 委員会 総務 委員会					2005	
6	しています。また、「国を挙げて環境問題に	取り組む	」といえば「国民全体」という意味になりま	新田 春夫	1940	男	書籍/8 言語	ドイツ語ことば の小怪	言語と文化の 日独比較		新田春 夫1著	大修館 書店	1993	
7	った被験者の方が課題にパズルやおもちゃに	取り組む	時間が長いというものである。パズルやおも	浦 光 博	1950	男	書籍/1 哲学	「心の質しさ」 を考える	「人間性」探究 の一助として		田村一 郎1編著	北樹出 版	2003	
8	全化のための計画を策定して財政の健全化に	取り組む	団体については、一定の行政水準を維持しな	自治省			白書/経済/地方 財政白書	地方財政白書 _昭和59年版				大蔵省 印刷局	1984	
9	うのでは、時間のむだになる。自給について	取り組む	べき課題は多い。なにかしたいと思う人にと	ジュディ ス・C・ ガラス, リチャー ド・E・ ネルソ ン, 那	1950	女	書籍/3 社会科学	友だちを自給 させないため にきみにでき ること			リチャー ド・E・ネ ルソン, ジュディ ス・C・ ガラス1 著;那波	アスベ クト	1997	

完了 F120%

3.2 JPwac-L2 の機能

Jpwac-L2 というのは、もともとSketch Engineと呼ばれるコーパス検索ツールに搭載されたJPwacと呼ばれるコーパスから抽出したデータである⁶。JpwacはWEB上の5万ページから収集し、整理された4億語規模のコーパスであるが、Jpwac-L2はそのデータから1億語を抽出し、その抽出したデータにいくつかの条件を設定し、その条件に合った文を抽出した用例コーパスである。Jpwac-L2に入れるための条件は、1) 1文の長さが25語以下、2) 20%以上の記号や数字を含まない、3) 日本語以外の表記を含まない、4) 句点で終わっている、5) 少なくとも一つの動詞、形容詞、形容動詞、あるいは助動詞を含むというものである。このコーパスは以上の条件を含んだ文で構成された約1300万語規模のコーパスである。

図3はJpwac-L2で検索を開始する画面である。検索する際には、ドロップダウンリスト機能により、使用するコーパスのレベルを設定することができる。レベル設定は旧日本語能力試験のLevel4(4級)～Level1(1級)、さらに上級のLevel0から選択でき、各レベルの文には当該レベルより上のレベルの語は含まないこと、当該レベルの語を10%以上含むことという条件が設定されている。検索結果はKOTONOHAと同様、KWIC表示されるが、チェックボックスの機能で各語のレベルやレンマ形、品詞などを表示させることや、キーワードや左右の表現をソートすることも可能である。KOTONOHAの機能と異なる点として、表示件数に限界がないこと、レンマでの検索も可能である点が挙げられる。図4は、「取り組む」という語を検索し、検索語の左側のテキストでソートを行った画面である。

図3 JPWac-L2の検索画面

図4 JPWac-L2の検索結果

Line	Word	Context	Link
80	word	このため、教育力向上に	取り組む
81	word	この間に	取り組む
82	word	そこで群馬弁護士会でこの問題に	取り組む
83	word	ですから、後は対処療法的に問題に	取り組む
84	word	けれども我々は信頼、開発の途中でこれらの問題に	取り組む
85	word	都市化と居住の問題に	取り組む
86	word	今後NU東京全体でVANの問題に	取り組む
87	word	私たちは今年中国問題に	取り組む
88	word	喫茶店をはしごしながら数学の演習問題に	取り組む
89	word	環境問題に	取り組む
90	word	京都のまちづくり問題に 取り組み 大文字ゴルフ場事件の勝利をきっかけに自然環境問題に	取り組む
91	word	正式名称は、部落問題に	取り組む
92	word	ただ、懸案の、PRTR 洗剤供給開始問題に	取り組む
93	word	一番良い防御は、常に難しい問題に	取り組む
94	word	収益金は、飢餓問題に	取り組む
95	word	問題の子どもに	取り組む
96	word	私は若い世代の宗教に	取り組む
97	word	この事件が国連が人種差別に	取り組む

4. 本学におけるデータ駆動型学習の取り組み

4.1 コロケーション学習の重要性

最後に実際に本学で取り組んでいるデータ駆動型学習について報告したい。日本語教育では漢字語彙学習の方法として様々な方法が提案されているが、その中でも、効果的な学習方法としてよく挙げられるものの一つに語の共起関係の学習（以下コロケーション学習⁷⁾）がある（加納 2000）。語彙学習の中で共起関係を学べば、意味的にまとまりのある語彙のネットワークを形成することができ、語彙量を効率的に増やすことができるというメリットがある。また、正確な意味の理解及び用法の理解をしながら、語彙学習をするためには、コロケーション学習が有効である。森田（1989）は語の意味や用法を明らかにするためには、明らかにしたい語の主体や対象、どのような文型で使用されるかなどを考える必要があることを述べている。つまり、学習者は語を単体で理解するのではなく、文法的な共起語、語彙的な共起語をとともに理解することで、意味の理解や用法の習得が効果的に行えるといえる。

本学においてもそれらの効果に注目し、コロケーション学習を漢字語彙学習の中に積極的に取り入れてきた。寺嶋他（2009）は教材の中で学習する語と共起語を提示し、学習者がそれらの語で短文を完成するという学習方法を実践している。その結果、コロケーション学習は中級の段階ではコロケーションに対する意識により大きな変化をもたらし、内容がより複雑になる上級Ⅰの段階では表現力を向上させる手段になると学習者により強く認識されていることがわかった。しかし、実際に提示されている共起語は限定されているという問題があるため、学習者自身が共起表現を調べられるように指導することが課題となっていた。そこで、2008 年秋学期から上級Ⅱにおいて学習者自身がコーパスを用いて共起する表現を調べ、短文を作るという漢字語彙学習方法を導入している。

4.2 本学におけるデータ駆動型学習の概要

本学の日本語上級Ⅱという授業は、週に 4 クラス（95 分/1 クラス）、1 学期間（約 90 時間）をかけて学習を行う。上級Ⅱでは、一つのテーマに基づき、読解、発表、レポート作成を行っている。たとえば、2010 年春学期は環境問題がテーマの中心にあり、コンビニのレジ袋の有料化、コンビニの 24 時間営業、クロマグロ資源の問題などが具体的な学習テーマであった。2008 年秋以前はそのような活動を行うために、テーマに合った語彙を選定し、語とその読み方が書かれたリストを渡し、それを各自で自学するように指導していた。その背景には限られた指導時間の中で「読む」「書く」といったスキル指導に重点を置いているため、授業内で十分に語彙の指導ができないという現状があったからである。しかし、語彙リストだけの学習では使用する文脈やコロケーションに対する理解が進まないため、表現力がなかなか向上しないという問題があった。そのため、日本語上級Ⅱでは学習者自身で漢字語彙が効果的に学習できるような指導が必要であった。

現在は同様のリスト⁸⁾を渡した後、「かぼす」というデータ駆動型学習を支援するシステムを用い、漢字語彙学習を指導している。「かぼす」というのは、(Kanji and Vocabulary Learning Support tool) のことで、Excel をベースに開発したソフトウェアである。「かぼす」は、Excel、日本語のフォントが入っている Windows のパソコンであれば、使用することができる。「かぼす」は、使用者が自由にコンテンツデータを入力することができ、コンテンツデータを入力することでフラッシュカード、空欄補充問題などが自動的に生成される機能がある。上級Ⅱの学習者は、先に述べた KOTONOHA や JPwac-L2 を用いて、コロケーションの情報やそのコロケーションを用いた短文を「かぼす」に入力するというデータ駆動型学習を行えば、作成したデータで自動的に生成される練習問題で学習することが可能である⁹⁾。図 5 は「かぼす」を起動したときの画面で、データを入力するさいは Data Entry のボタンをクリックすると、図 6 の画面が表示される。また、「かぼす」の学習ツールで学ぶ場合は、図 5 の画面で選択すると、Flash Card（図 7）、Kanji Card（図 8）、Typing（図 9）といった画面が表示される。

2.3 でも述べたように、データ駆動型学習を導入するためには、学習の意義やコンコーダンスの使い方などのガイダンスが重要になる。そこで、授業の最初にガイダンスを行った。表 1 は、データ駆動型学習のガイダンスと「かぼす」の使い方のガイダンスの流れを示している。まずコーパスの紹介やコンコーダンスの使い方を説明した後、間違った表現を直す、類語表現の中でよく使われる表現を選ぶ、ある語とよく共起する表現を探す、品詞ごとによく共起する表

現を探すといったように、コーパスを使う目的を確認しながらタスクを行った。

そして、5回目のガイダンスが終わった後で、手渡した語彙リストから6語を選び、図10のような形式の紙を学習者に渡した。そして、JPWac-L2やKOTONOHAといったコンコーダンスー、オンライン辞書などで、必要な情報を調べて紙に書くように指導した。紙に書くのは、品詞、英語での意味、アクセントの番号、学ぶ語の前に強く共起する語、後ろに強く共起する語、そして、それらの共起語を使った短文である。短文を作成するという指導は寺嶋他(2009)での方法と同じで、本学の学習者はそのような学習を中級から続けている。しかし、それまでの学習方法は、与えられたコーンケーションに基づき、短文を作成するという方法であった。

6回目のガイダンスでは、「かぼす」を配布し、使い方のガイダンスとデータ入力の練習を行った。「かぼす」にデータを入力するさいは図6の画面で行うが、そこに入力する情報は図10で練習したものと同じである。そして、ガイダンス終了時には、各クラスで語彙リストに基づき、データを作成するように学習者に指示を出した。ただし、1回当たりの語は60語から70語ほどあるため、一人の学習者で全てを作るのは負担がある。そのため、各クラスで担当を決めてデータを作成し、共有するようにした。また、学習者が作成したデータには問題も見られるため、クラスで共有する前に教師は一人一人の学習者にフィードバックを行った。しかし、この方法では学習者によって作成するデータが異なってしまう、指導する語が異なるという問題が生じるため、共通に指導したい語がある場合は図10の形式の紙を配り、指導を行った。

表1 漢字語彙学習の流れ

テーマ	回数	内容
データ駆動型学習のためのガイダンス	1	コーパス及びコンコーダンスーの使い方の説明 間違った表現を確認し、どのような表現が適切かをコーパスで調べる 例) <u>目標を完成するために</u> 、努力している。
	2	類似した表現を示し、どちらの表現がよく使用されるかをコーパスで調べる 例) <u>割合が増える。</u> ／ <u>割合が高くなる。</u>
	3	〈NT〉の部分でよく使用される共起表現を調べる 例) 〈NT〉を身に付ける
	4	品詞ごとでよく使用される表現を調べる 例) <形容詞>+取り組み、<名詞>+取り組み、<名詞+助詞>+取り組み <副詞>+取り組み
かぼすの使い方のためのガイダンス	5	紙でのコンテンツ作成(図10) かぼすの機能とデータの入力方法を学ぶ
	6	「かぼす」を使っでのコンテンツ作成(図6)

図5 起動画面

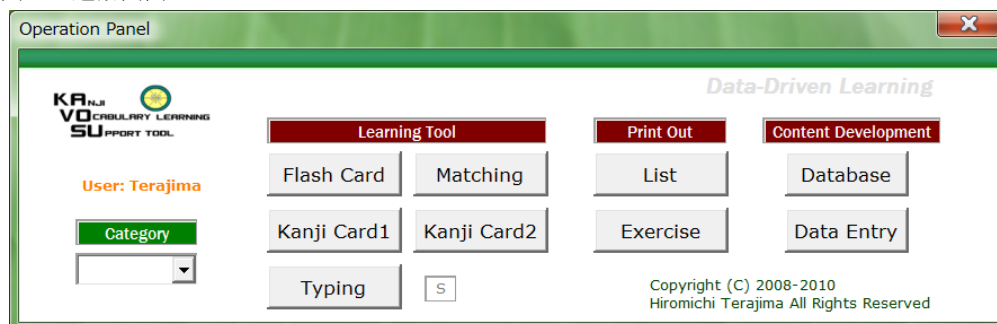


図6 データの入力画面

The 'Data Entry' window is a form for inputting word data. It includes fields for 'Word' (危機), 'Learning' (radio buttons), 'ID' (68), 'Reading' (きき), 'Accent' (1), 'Collocation' (財政, 食糧, 経済, 石油), 'Example' (何十年の平和が続いてきたため、我々は「危機感」を忘れてしまった。), 'Part of Speech' (noun), and 'Category' (AD2-2). There are also buttons for navigation and search.

図7 フラッシュカード

The 'Flash Card' window displays a word card for '増大' (increase). It includes a timer (7 seconds), a 'Word' field, and an 'Answer' field. The user is prompted to check the reading and meaning of the word. There are buttons for 'Next' and 'Answer'.

図8 漢字カード

The 'Kanji Card2' window displays a sentence: '環境保護のため、ゴミの[]は適切に行われている。' (For environmental protection, the [] of garbage is appropriately handled). The user is prompted to choose two appropriate Kanji from the options: 大 (big), 表 (surface), 処 (place), and 理 (reason). There are buttons for 'Check', 'Answer', 'Next', and 'Score'.

図9 タイピング

The 'Typing' window displays a sentence: 'もし電機製品を買うなら、新聞[]を見たほうが いいと思う。' (If you buy electronic products, I think it's better to see [] in the newspaper). The user is prompted to type the appropriate word in the Input Field. There are buttons for 'Check', 'Answer', 'Next', and 'Score'.

図10 学習者によるデータ駆動型学習の例

AD2-2

名前

6	依存 ^{negative} いぞん noun+する (type III) dependance	- アルコール, ネット, たくさん, 相互, 輸入, 公債 + 症, する, 度, 性的, 的, 関係, 型	0
		日本は石油がないので、外国からの事前に依存している。 ストレスがたま、た人はアルコール依存症に陥りやすい。	
8	危機 きき noun crisis	- 重大な, 倒産の, 経済, 金融, 石油 + 管理, 感的, 的, に, 渉及, する, を, 乗り越える	12
		経済危機を乗り越えるために、みんなの努力が必要だ。 その会社は重大な危機に陥っている。	
9	被害 ひがい noun damage	- 地震の, 直接, 健康, 拉致, 犯罪 + が出る, がある, に, 遭う, を, 与える, を, 受ける者	1
		地震の被害者は援助をまだもらっていない。 その水質汚染はみんなに健康被害を与える。	
13	絶滅 ぜつめつ noun+する (type III) extinction	- 種の, 大量, 恐竜の, 生物 + する, の, 危機, 危惧, 種, 収容所	0
		人間の行動が原因で絶滅する生物が多い。 生物絶滅を防ぐために、環境保護が必要だ。	
5	見込む みこむ vs (type I) expect, anticipate	- 増加を, 将来を, 利益を + とし, 客, 企業, 必要	02
		来年、その企業は利益の増加を見込んでいる。 不景気なので、多客の利益を見込む企業が減少している。	
18	留まる とどまる v1 (type II) stay, keep, fix, fasten, stop	- 目に, 思い, 歩, 踏み, 固に + 場合, と, ころ, 的, べき	3
		彼は留学することを思い留まった。 彼の月宛白は留まることを知らない。	

4.3 実践したデータ駆動型学習の考察

今回のデータ駆動型学習の特徴として考えられるのは、よく使用される共起表現を自由に探すという点である。共起表現を探すというのは、中條他 (2009) でも見られるが、それらの教材では、「～への対応」、「に対応した～」 「援助が～」といったように構造が指定され、その構造に当てはまる語彙的コロケーションを探すようになっている。そのような学習は、今回のデータ駆動型学習のガイダンス (3 回目) で行っており、今回実践したデータ駆動型学習は、より自由度と難易度が高いものである。学習者の中には、複合語の表現ばかりを選ぶ者や途中でデータを探すのを諦め、コンコーダンサーで表示された最初の方に現れる共起表現を選び、最後まで画面をスクロールしない者も見られた。そのような場合はフィードバックを行い、より適切だと思われるコロケーションを取り上げるよう導いた。

二つ目の特徴は調べたコロケーションを用い、短文を完成する点である。効果的に言語習得を行うためには、理解と産出が必要である。特に上級段階になると、文章表現の指導が強調されるが、まずは短文で産出できる能力を身に付けることが必要である。そのような理由から、短文を完成させる指導というのは、中級から一貫して行ってきた。そのため、今回の授業でコロケーションを使い、短文を完成するという点自体は、それほど抵抗がないようであった。しかし、寺嶋他（2009）が報告している例と同様に、検索した以外の部分で不自然なコロケーションを生成したり、コロケーション以外の部分で文法的な間違いをおかししたりする例もみられた。他にも学習者自身がコロケーションの候補としてあげた語を短文完成の中で全く使わないという例、前後にあげたコロケーションを無理に使用し、不自然な表現になっている例も見られたため、フィードバックを行った。

三つ目の特徴は、学習者は語彙情報を同じ形式でリストのようにまとめていくという点である。中條他（2009）では、たとえば、「対応」という語を取り上げ、様々な質問に答えることで、理解を深めるという方法を取っている。しかし、実際の教育現場では、一つの語を学ぶのに、それほど多くの時間をかけられないのが現実である。そのため、できるだけシンプルな形式で語彙情報をまとめるようにしている。しかし、それだけでは単調な学習になるという問題も考えられる。そのため、「かぼす」のフラッシュカード、空欄補充問題などといった機能で学習できるようにし、理解を深められるような環境を整えたが、実際にはそのような学習ツールを用いず、「かぼす」で作成したデータをリストとして印刷し、リスト化された紙に様々な情報を書きこみながら学んでいる学習者も多く見られた。

2.3 で述べたように、コーパスの難易度に変化を感じている学習者もいたが、そのような学習者にはコンコーダンスーに表示された文の内容を全て理解するのではなく、検索語とその前後の語の共起関係と意味に注目し、理解するよう指導した。また、コンコーダンスーの使い方がわからない学習者や学習に対する態度が前向きではない学習者もいたが、練習と説明を重ねるにつれ、そのような学習者は減少していった。学習者の中には、最後までデータ駆動型学習に馴染めないような学習者もいたが、作成するデータの質が向上していった学習者、レポートにもその効果が表れている学習者も見られたため、今後詳しく分析し、指導方法を改善していく必要がある。

5. まとめ

以上、本稿では、日本語教育でデータ駆動型学習を行うための学習方法を整理し、実際に本学で行っているデータ駆動型学習の取り組みを報告した。

データ駆動型学習はオーセンティックな言語に大量に触れることができ、学習者の疑問を解決するための新たなストラテジーにもなると考えられる。また、言語現象に対する分析力を高められる可能性があるだけでなく、ブルーナーが提唱した発見学習の効果と同じように、法則を発見しようという身構えた心を育成できる、学習への動機を高められる、記憶が長く保持されるといった効果も得られる可能性があり、さらに実践的な研究を進める必要がある。しかし、データ駆動型学習は、いくつかの問題も見られるため、授業で導入するさいには注意が必要である。今後は、継続的にデータ駆動型学習の実践と改善を行い、データ駆動型学習の可能性を探っていきたい。

注

1. 国立国語研究所の文部科学省科学研究費特定領域研究の一部として、現代日本語書き言葉均衡コーパスの構築が進められている。その中の一つの組織である「作文支援システム班」がデータ駆動型学習を可能にするシステム「なつめ」を開発しており、それによって作文支援を行う研究プロジェクトがある。
2. 国立国語研究所 『KOTONOHA』 <http://www.kotonoha.gr.jp/demo/>
3. 川村 よし子（東京国際大学）, クリスティナ・フメリヤク・寒川（リュブリャーナ大学）『JpWac-L2』
<http://nl.ijs.si/jaslo/cqp/>
4. KWIC 表示とは、Key Word in Context の略で中央に検索した語が並び、その左右に文脈を出現させる機能である。左右の共起語を調べるために使用される機能である。
5. 他にも、東京工業大学留学生センター仁科研究室で開発されている「なつめ」がある。現在、「なつめ」では毎日新

聞や青空文庫などのテキストから検索が可能であるが、今後国立国語研究所の「現代日本語書き言葉均衡コーパス」から検索ができる予定である。

6. Jpwcについてはスルダノヴィッチ・エリャヴェツ・イレナ他 (2008) が詳しく報告している。
7. コロケーションの定義は野田 (2007) に従う。
8. 語彙リストは漢字語彙テストの数に合わせ、1回目～5回目までに分けられている。1回あたりに60語～70語の漢字語彙が含まれている。
9. 漢字語彙を紙媒体で学習したいという学習者のために、入力した情報をリストとしてプリントできる機能、練習問題がプリントアウトできる機能もある。

参考文献

- J・S・ブルーナー (著), 鈴木祥蔵, 佐藤三郎 (訳) (1963) 『教育の過程』 岩波書店
- Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 293-313). Cambridge: Cambridge University Press
- Noguchi, J. (2004). A Genre Analysis and Mini-Corpora Approach to Support Professional Writing by Nonnative English Speakers 『英語コーパス研究』 第11号, 英語コーパス学会, 101-110
- Tribble, C. and Johns, G. (1997). *Concordances in the Classroom*, Houston: Athelstan
- 石川慎一郎 (2009) 「日本語基本研究における非統制型・統制型・媒介型 Web as Corpus の可能性—言語コーパスにおける基本語頻度の安定性について」 『特定領域研究 日本語コーパス 平成20年度公開ワークショップサテライトセッション予稿集』 特定領域研究日本語コーパス総括班, 29-38
- 石川慎一郎 (2008) 「言語コーパスとしての WWW—広がる可能性—」 『日本語学』 27 巻 2 号, 明治書院, 10-21
- 岩井千春 (2009) 「コーパス言語学を応用した ESP 教育—発見学習に基づいた学習者自律」 『大阪府立大学紀要, 人文・社会科学』 57 巻, 1-12
- 梅咲敦子 (2008) 「英語教育とコーパス—問題解決型学習と自律した英語使用のためのコーパス利用—」 『日本英語コミュニケーション学会紀要』 17 巻 1 号, 115-117
- 荻野綱男 (2008) 「WWW をコーパスとして利用する研究—文系と理系の観点から—」 『日本語学』 27 巻 2 号, 明治書院, 4-9
- 加納千恵子 (2000) 「中上級学習者に対する漢字語彙教育の方法」 『筑波大学留学生センター日本語教育論集』 15, 35-46
- 川村よし子、クリスティナ・フメリヤク・寒川 (2010) 「Web コーパスを活用したレベル別例文検索システムの開発と評価」 『ヨーロッパ日本語教育』 Vol. 14, pp. 226-233
- スルダノヴィッチ・エリャヴェツ・イレナ, 仁科喜久子 (2008) 「コーパス検索ツール Sketch Engine の日本語版とその利用法」 『日本語科学』 第23号, 国書刊行会, 59-79
- 投野由紀夫 (2003) 「コーパスを英語教育に生かす」 『英語コーパス研究』 第10号, 英語コーパス学会, 249-264
- 野田尚史 (2007) 「文法的なコロケーションと意味的なコロケーション」 『日本語学』 26-11, 明治書院, 18-27
- 中條清美, 西垣知佳子, 内山将夫, 原田康也, 山敦史 (2005) 「日英パラレルコーパスを活用した英語語彙指導の試み」 『日本大学生産工学部研究報告 B』 第38巻, 日本大学生産工学部生産工学研究所, 17-37
- 中條清美, 田辺和子, 木下謙朗, 三橋麻子, 西垣知佳 (2009) 「コーパスを活用した日本語教材作成の試み」 『日本大学生産工学部研究報告 B』 第42巻, 日本大学生産工学部生産工学研究所, 43-52
- 寺嶋弘道, 小林潔子 (2009) 「コロケーション情報を用いた漢字語彙学習の試み」 『ポリグロシア』 17 巻, 139-149
- 姫野昌子 (2006) 「学習者のためのコロケーション辞典 - 『日本語表現活用辞典』の作成に際して - 」 『日本語学』 25-8, 明治書院, 40-5
- 前川喜久雄, 山崎誠 (2009) 「現代日本語書き言葉均衡コーパス」 『国文学解釈と鑑賞』 74 巻 1 号, 至文堂, 15-25
- 水越敏行 (1975) 『発見学習入門』 明治図書出版

日本語教育におけるコーパスの応用
ーデータ駆動型学習とその実践方法の考察ー

森田良行（1989）『基礎日本語辞典』角川書店